Identification of Southern French accent based on suprasegmental elements

Annelise Coquillon

Identification of Southern French accent based on suprasegmental elements

Annelise COQUILLON
Laboratoire parole et langage, Institut de Phonétique / CNRS ESA 6057
Université de Provence (Centre d'Aix), FRANCE
E-Mail: annelise.coquillon@lpl.univ-aix.fr

ABSTRACT

This work focuses on the identification of two different varieties of French on the basis of suprasegmental elements. More precisely, we will show that a perceptual discrimination between Southern French (spoken in the Marseilles area) and a more neutral French (which can be assimilated to standard French) is possible using prosodic cues alone. For this, we used a low-pass filter adapted to each speaker. As these results show the importance of the suprasegmental elements of Marseilles French, we investigated a preliminary research of the prosodic parameters of the corpus. This first investigation enabled us to draw certain trends of this regional accent which will further be discussed.

1. INTRODUCTION

Several studies showed that a speech signal reduced to its prosodic components (filtered utterances) contains the required information for the identification of the language it conveys ([1], [2]). Moreover, it has been observed [3] that the prosodic differences involved in the dialectal variants of the same language can sometimes be more prominent than the ones that distinguish two different idioms. Although it is generally agreed that the role of prosody in the identification of dialects or regional variants is of great importance, very few experimental studies have investigated on this question for French. Therefore, we propose to look further into this field, where the current state of knowledge remains very limited.

The aim of this paper is to widen the state of knowledge on Southern French prosody, especially the one spoken around Marseilles. For this purpose, we propose to compare this dialectal variation with a more "neutral" French accent, i.e. containing as little dialectal specificities as possible. This latter is commonly referred to as "standard", though we would rather call it "neutral" French.

2. SPEECH MATERIAL

Regarding the nature of the problematics of this work, i.e. the comparison between the French spoken in Marseilles and a more normative French, we choose to start with laboratory speech, in the form of informal dialogues read by pairs of two speakers. This enables a direct comparison between the two varieties, as well as to control most of the variables. Another advantage of interpreted dialogues is to lessen the risk of a neutralization of the dialectal characteristics, as it is often the case in read speech [4]. The utterances were therefore taken from real spontaneous conversations. The result could be better considered as semi-spontaneous speech than as read speech. We subsequently proceeded to the recording of an interview of the speakers so as to have a more spontaneous communication, allowing us to validate the hypothesis obtained on the read corpora. This corpus was recorded in an anechoic chamber, by sets of two persons each equipped with a head-microphone. The speakers were first selected on linguistic criteria. Indeed, it has been shown that the Southern French accent comes mainly from the Provencal substratum [5]. Therefore, we limited the geographical origins of the southerners to the former linguistic borders of Maritime Provencal, which was spoken in the area of Marseilles. As for the "control" group, they were chosen to represent "neutral" French, i.e. not marked dialectally. They are all native of north of France, so that their accent has never been influenced by the southern one.

This of course needed to be estimated by perceptual tests. We selected from the read dialogues the three following sentences for every speaker, so that the lexical information would not influence the perception of the regional accent.

- Ma foi, toujours pareil... Je fais le manœuvre de temps en temps tout ça. Et toi, toujours dans les études ? ("Oh well, as always… I'm doing unskilled work from time to time. How about you, still studying?")
- Ben, la dernière nouvelle, pardi ! Je vais être tonton / tata. ("well, the last piece of news, of course! I'm going to be an uncle / aunt.")
- Dis, faudrait qu'on se fasse un bon aï oli[1] un de ces quatre. Tiens, qu'est ce que tu fais samedi ? ("Say, we've got to have a good aï oli one of these days. Look, what are you doing next saturday?")

This data, presenting 6 speakers for each regional group, was subsequently submitted to an auditory assembly of 35 persons who were asked first to determine if the speakers originated from the region of Marseilles or not, and second to specify whether their regional accent was more or less

[1] "Aï oli" is a regional word, taken from the Provençal language. It refers to a local dish.

marked. The subjects of this test, 16 men and 19 women, were aged from 24 to 58 years old (35 in average). 22 of them live in the Marseilles area, amongst which 16 natives. However, no distinction could be made between the answers of the listeners on account of their regional origin. This previous validation of the corpus permitted to select five speakers for the southerners (four males and one female, aged from 25 to 40 years) on the basis of their strong Marseilles accent and four (three males and one female, aged from 28 to 34 years) for the pilot group on account of the regional neutrality of their accent. They are considered as representative of their regional group. (see the general scores in table 1).

One may find interesting to remark the important variability among the answers concerning the degree of accent of a speaker. For instance, one of the subjects (M6) was evaluated as not native of Marseilles by 14% of the listeners, whereas 17% of them judged his Marseilles accent as being very strong. This shows how important it is, at least when working on regional variations, to submit a corpus to a group of listeners. Indeed, the notions of norm (as a standard for a language), or "regionality" or again the degree of an accent are subjective and should not, to our opinion, be assumed by the researcher alone.

3. DISCRIMINATION TEST

3.1. Method

In order to evaluate the relative importance of the prosodic cues in regional identification, we elaborated a perception test where the listeners would not be able to rely on the lexical or segmental elements. A number of spontaneous (interview) as well as semi-spontaneous (interpreted dialogue) speech files were selected from the corpus (mean duration of 19:18 seconds). Therefore, the same speaker is presented twice, but in two different speaking styles. The selection was reduced to its prosodic expression, where the segmental characteristics (phonemic and lexical) are annihilated or at least limited, by means of a low pass filter, using the PRAAT program [6]. The levels of filtering vary with the speaker, ranging from 350 to 500 Hertz (the highest corresponding to women). These were perceptually validated by a set of ten listeners, asked to determinate if the segmental elements could not be heard anymore. The outcome could be compared to a humming.

We are aware that this method does not allow to know accurately which properties of the signal are eliminated or preserved [1]. However, most segmental information should be eliminated at this level, as it is mainly contained in the higher formants of speech, and pitch should be preserved as we took care that it would never rise higher than the cut frequencies we choose for each of the speaker.

An auditory group of 25 subjects was then required to determine from this filtered corpus the regional origin of the speakers using the prosodic cues that remained. The subjects, 10 females and 15 males, were at that time aged from 24 to 58 years (36 in average), and 17 of them lived around Marseilles (but only 10 were born there). The perceptual task thus consisted in determining whether the speakers originated from Marseilles or not. They were also asked to evaluate whether they heard read or spontaneous speech, so as to establish the naturalness of the read dialogues. For this reason, we tried to avoid to have two speakers say the same sentence (for the read part of this test) so that the listeners would not get used to the utterance and therefore detect that is was read.

3.2. Results

Overall, the results were highly above chance level. The average of correct identification of the speakers' regional origin was 75%, which is close to the scores obtained in a similar task for different languages (86% [2] for bilinguals between French and Spanish using low-pass filtering; 66,9% [1] between English, French and Japanese using resynthesis). The listeners of the test found the task quite difficult, though individual good answers vary from 55 to 100% (only one of them made no mistake). Most said to have relied mainly on intonation, but rhythm or articulation rate were also some of the cues mentionned. We expected the Marseilles listeners to be better performers for this discrimination task, as bilinguals are for similar tasks concerning two languages. Nevertheless, there seemed to be no direct relation between the scores of the subjects and their geographical origin. Indeed, a factorial analysis of variance (Anova) was not significant at the 0,05 level [$F(1,478)=2,272$; $p=0,13$].

Individual results (by speaker) range from 58% to 92% in average of good classification into regional groups. They are shown in table 1 below, in parallel with the representativeness found in the validation test, as mentioned earlier. The speakers from Marseilles are identified by "M" and the "Standard" ones by "S". They were previously classified from the better (M1 or S1, see also figure 1) to the worst identified within each group, though in this table they appear from the most to the less representative of each region.

speaker	mean identification	representativeness
M1	92 %	88 %
S1	88 %	85 %
M4	73 %	73 %
S3	67 %	66 %
S2	69 %	55 %
M2	88 %	54 %
M3	85 %	43 %
M5	69 %	40 %
S4	67 %	37 %
M6	58 %	20 %

Table 1: Representativeness of speakers crossed with mean identification in filtered tests.

The results of the two tests seem to be somewhat related. Indeed, the two best (M1, S1) and the two worst (S4, M6) recognized speakers correspond respectively to the two most and the two less representative of their regional group.

Detailed results, i.e. by speaker and speaking style, show greater variation as the mean of correct identification scores fluctuate between 42 to 96%. However, only one score was below 50%, as shown in figure1.

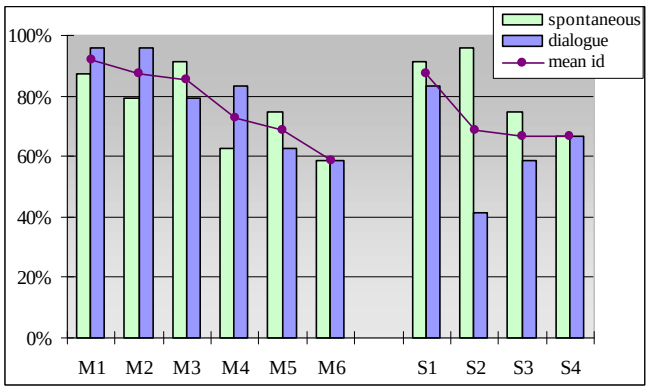

Figure 1: Mean identification scores per speaker detailed in spontaneous and dialogue speaking style

The order in which the speakers appear in the figure above is, within each dialectal group, from the best identified, in average, to the worst (see dotted line "mean id"). Overall, we can see that most speakers were better or evenly identified in interview situation than while performing read dialogues. The three exceptions are to be found in the Marseilles group. Still, note that the listeners did recognize that they were in a reading situation by 53% (M2) and 74% (M1 and M4) of them. Also, it is surprising that speaker S2 who obtained the worst identification score in read dialogue (42%) also obtained one of the best result in spontaneous speech (96%), with a difference in ratio of 27% between the two scores. Though 95% of the subjects of the test perceived that he was in a reading situation, this difference could not yet be explained, and will require a detailed prosodic analysis of the data. Aside from this speaker, the proportional difference between good identification in dialogues and in spontaneous speech are situated between 0 to 10% (8% in mean), which again isn't sufficient to favor a speech style for this type of test. What's more, the factorial Anova revealed that the situation of speech with the identification in filtered test is not a significant parameter : [F(1,478)= 2,272 ; p = 0,13].

4. FIRST INVESTIGATIONS

These results confirm that the prosodic elements of the Marseilles French are sufficiently specific to be distinguished from a normative French. Therefore, we will present here a first analysis of some prosodic differences between these two varieties. First, we will produce the speech rate of the speakers, taken on the passages selected for the perceptual test presented in chapter 3 so as to see if this parameter could have contributed to the discrimination between the dialects. Secondly, we will give some global tonal characteristics on the entire data, and then present the selection that was made on the corpus for a more local analysis.

4.1. Speech rate

As the speech rate is one of the cues listeners said to have used in this filtered test, we calculated each speaker's articulation rate on the presented data. It corresponds to the mean number of syllables spoken per second (syll/s) [7]. We will present here only the speech rates which include pauses, knowing that similar results were found with data without pauses. Individual results showed some dispersion between the speakers (3,69 to 6,2 syll/s). In average, all show a higher articulation rate in read dialogues than in interview (5,61 against 4,89 syll/s). Also, the speakers from Marseilles tend to speak faster than their counterparts (mean 5,5 against 4,88 syll/s) [F(1,478)=163,048 ; p<0,0001]. Anyhow, no relationship between individual speech rate and the discrimination scores obtained in the filtered test could be made.

4.2. Global characteristics

As the passages used in the test mentioned above are samples of the corpus, we performed further analyses on the whole data available (entire dialogues and interview), without the two speakers who obtained the worst mean scores of identification, as they also correspond to the speakers estimated as the less representative (M6 and S4). We estimated here the tonal amplitude of the speakers, on the basis of tonal targets, automatically detected by MES [8]. We then calculated the pitch range as the mean of the highest targets divided by the mean of the lowest one [9], given on an ERB scale. We also mention for information the mean of all the targets, its standard deviation, and the maximum and minimum target values, as shown in table 2 below.

Speaker	f0 range	mean (ERB)	s.d.	Max (ERB)	Min (ERB)
S1	1,74	3,97	0,56	6,62	2,48
S3	1,74	5,08	0,73	8,31	3,29
M5	1,89	4,18	0,76	6,75	2,40
S2	1,98	4,12	0,69	8,21	2,64
M4	2,12	4,56	0,90	7,49	2,51
M1	2,13	4,70	0,88	7,07	2,54
M2	2,15	3,71	0,70	6,95	2,39
M3	2,78	5,43	1,71	9,68	2,52

Table 2: Pitch measures in ERB : f0 range, mean of f0 targets, maximum and minimum values.

In table 2, speakers are sorted by f0 range, from the narrowest to the wider range. As we can see, the speakers from Marseilles tend to use a wider range than the others do, except for M5. This might be one of the reason why the latter was not estimated as having a very strong accent (40%, as shown in table 1), and was neither very well identified through the filtered perception test (64%).

4.3. Local parameters

4.3.1 Locating relevant passages

In order to locate in the Marseilles corpus the most relevant passages for the analysis, we submitted it to a group of

phoneticians specialists. They were asked to underline in the corresponding text where the southern accent seemed, at the prosodic level, the most pronounced. Thus, we recorded the parts underlined at least by three out of the six phoneticians for each of the speakers. We then selected the noted passages shared by at least three out of the six speakers. Overall, 69 passages were underlined commonly for at least three speakers.

4.3.2. Results

The selection of the passages by the underlining test concerns particularly intonative unit endings (59 out of the 69 underlinings, which represent 86% of them). These include 20 continuative units, 13 questions and 26 terminal utterances. In French, word or intonative unit endings carry indeed important prosodic information, as this language presents a phrase final accent where main stress is realized, characterized by lengthening and located at prosodic boundaries. Therefore, we suggest that these underlined passages carry important prosodic information specific to the region of Marseilles, which we will further submit to comparison with the same passages for the standard speakers.

Also, within the 69 passages retained, 30 present a word final schwa, and 34 at least a nasal vowel (45 occurrences of nasals in all, as some words or group of words can count more than one nasal). It has been shown that these two elements are the most pregnant southern dialectal marks. Indeed, nasal vowels are often longer in Southern French than in other variants of French, as they are most of the time composed of 3 segments (as opposed to only one): They begin with an oral vowel, which is then often nasalized, and end with a consonantal segment [10]. Besides, the final schwa, called "mute e" in general French, is almost always pronounced in the South of France, which consequently introduces an unaccented final syllable, rare in French, and therefore a particular rhythmic pattern [11]. These might be a starting point to the cues used in the discrimination task.

5. DISCUSSION AND CONCLUSION

The results of the perceptual experiment presented in this paper confirm the hypothesis that the French spoken in Marseilles can be distinguished from a more regionally neutral French on the basis of prosodic cues alone. It also appears that spontaneous speech (interview) tend to be a somewhat more suitable discourse type for this kind of task than read dialogues, though this could not be statistically accounted for in this study. Even though it wasn't possible with the presented data to show which prosodic parameters contributed to a better identification, a first analysis allowed us to draw certain trends of this dialect. Indeed, it globally emerged that the Marseilles speakers tend to speak faster and to use a wider pitch range than the pilot group. A selection of the most relevant passages for the study of the prosodic specificities of the corpus will be a starting point for a more thorough investigation.

REFERENCES

[1] F. Ramus and J.Mehler "Language identification with suprasegmental cues: A study based on speech resynthesis", *Journal of the acoustical society of America,* 105(1), pp 512-521, 1999.

[2] M. Le Besnerais, "Reconnaissance par des locuteurs monolingues et bilingues de l'espagnol et du français à partir de productions filtrées ", Proceedings of the XXIèmes Journées d'Études sur la Parole, pp.47-50, 1996.

[3] D.J. Hirst and A. Di Cristo, *Intonation systems: a survey of twenty languages*, Cambridge University Press, 1998.

[4] M. Le Douaron, "contribution à l'étude socio-linguistique et phonétique des parlers méridionaux", PhD Thesis in phonetics, University of Provence, 1983.

[5] J. Seguy "Le français parlé à Toulouse", ed. Privat, 1978.

[6] PRAAT, phonetic software created by P. Boersma and D. Weenink, www.praat.org.

[7] D. Duez and Y. Nishinuma: "Vitesse d'élocution et durée des syllabes et de leurs constituants en français parlé". In Institut de Phonétique d'Aix et UA 261 du CNRS , pp.157-180, 1987.

[8] R. Espesser : "Mes: un environnement de traitement du signal", Actes des XXIèmes Journées d'Etudes sur la Parole, pp 447, 1996.

[9] D. Patterson, "A linguistic approach to pitch range modelling", PhD thesis, University of Edinburgh, pp. 199, 2000.

[10] S. Clairet, "Caractérisation des voyelles nasales en français méridional : essai méthodologique pour une analyse aérodynamique et acoustique", DEA in phonetics, University of Provence, 1997.

[11] A. Coquillon : "Marseillais et toulousains gèrent-ils différemment leurs pieds ? Caractéristiques prosodiques du schwa dans les parlers méridionaux" Proceedings of the XXIIIth Journées d'Études sur la Parole, pp.89-92, 2000.